Continue with regression

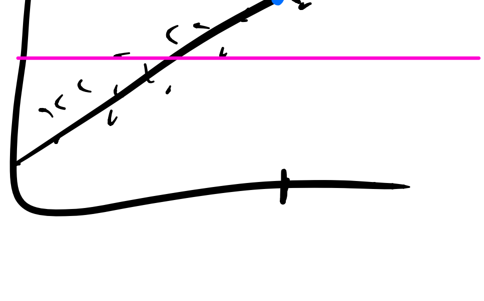
Some data (x, y)

$$\text{lm}(y \sim x)$$

$$\neq$$

$$\text{lm}(x \sim y)$$

Errors from predictions



rms error = rms of pred error

pred error = actual − pred
error

If we just guessed the average of y for every x then RMS error $= SD_y$

For regression

$$\text{rms error} = \sqrt{1-r^2} \cdot SD_y$$

since $-1 \le r \le 1$

$$\Rightarrow 0 \le \sqrt{1-r^2} \le 1$$

if $r=1 \Rightarrow \sqrt{1-r^2} = 0$

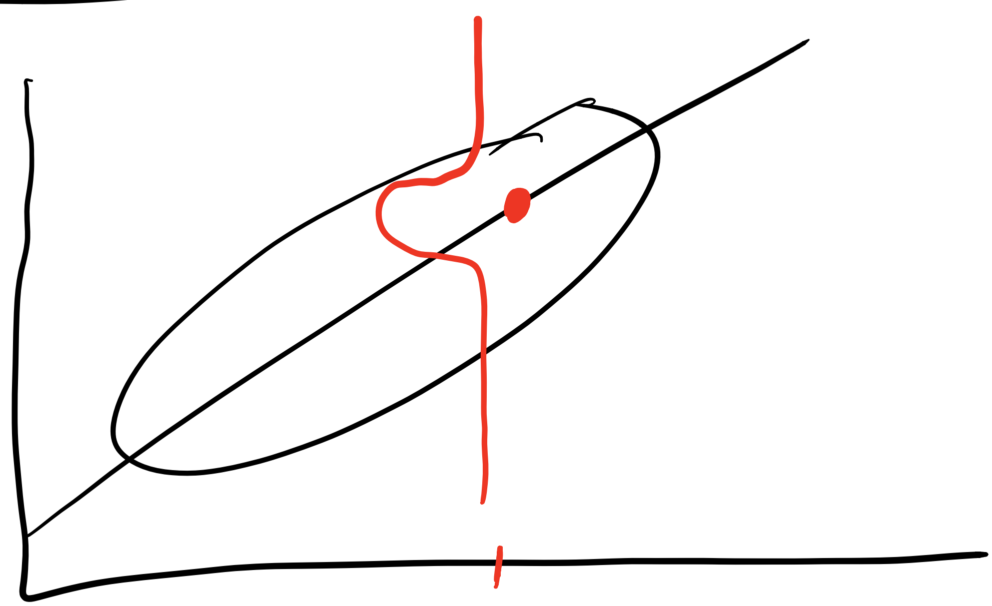
perfectly linear relationship
error $= 0$

if $r=0 \Rightarrow \sqrt{1-r^2}=1$

RMS reg $= SD_y$

no better off than guessing the avg of y.

The strips of normality



The equation of the line

$$y = mx + b$$

For the lin. regression

$$m = r \cdot \frac{SD_y}{SD_x}$$

$$b = avg_y - m \cdot avg_x$$

If you recall last time

$$y_p = \left(\frac{x - avg_x}{sd_x}\right) \cdot r \cdot SD_y + avg_y$$

$$= x \cdot r \cdot \frac{SD_y}{SD_x} + avg_y - avg_x \cdot r \cdot \frac{SD_y}{SD_x}$$

$$= x \cdot m + avg_y - avg_x \cdot m$$

$$= mx + b$$